

修士学位論文

論文題目 ユーザの体験情報を用いた観光支援シス

テム

氏
ふ
り
が
な
名

潘
はん
健太
けんた

専攻

情報学専攻

指導教授

北山 大輔 深教授

修了年月(西暦)

2020年3月

工学院大学大学院

要旨

目次

第 1 章	序論	1
第 2 章	本研究のアプローチ	4
2.1	観光支援システム	4
2.2	関連研究	5
第 3 章	ユーザの既訪問スポットの位置付けに基づく未訪問スポットの説明手法	8
3.1	概要	8
3.2	未訪問スポットと既訪問スポットの対応付け	9
3.3	対応関係を表す特徴語抽出評価	11
3.4	観光スポットの対応付けに関する評価	13
3.5	未訪問スポット説明の有効性評価	15
第 4 章	地図上における未訪問スポットの説明性向上のための観光スポットの対応 関係可視化手法	17
4.1	概要	17
4.2	観光スポット間の対応関係を用いた可視化	17
4.3	対応関係の可視化評価	18
第 5 章	考察	24
第 6 章	結論	25
	謝辞	26
	本研究に関する発表	27
	その他の発表実績	28
	受賞	29

図目次

1.1	レビューサイト上で観光スポットに付与される情報例	2
2.1	観光支援システムの流れ	5
3.1	プロトタイプシステム	9
4.1	既訪問スポットの経度座標計算の例	19
4.2	位置関係 (<i>Position_Map</i>)	20
4.3	位置関係 + 線 (<i>Line_Map</i>)	20
4.4	表形式 (<i>Table_Map</i>)	20

表目次

3.1	既訪問スポット集合と未訪問スポット集合	11
3.2	対応付け結果の例	11
3.3	対応関係を表す特徴語の実験結果	12
3.4	対応付けに関する実験結果	14
3.5	既訪問スポットのカテゴリの同異による評価割合	15
3.6	システム a の方が良いと回答した被験者の自由記述	16
3.7	システム b の方が良いと回答した被験者の自由記述	16
4.1	行き先の決定しやすさの回答数割合	21
4.2	<i>Position_Map</i> と <i>Line_Map</i> の訪問履歴との関係の回答数割合	21
4.3	<i>Line_Map</i> と <i>Table_Map</i> の訪問履歴との関係の回答数割合	21
4.4	<i>Position_Map</i> と <i>Line_Map</i> の決定しやすさと関係わかりやすさについての回答数	22
4.5	訪問先決定までにかかった体感時間の回答数割合	22
4.6	訪問先決定までにかかった労力の回答数割合	22

第 1 章

序論

一般に、ユーザは観光旅行する前に Web 上の情報やガイドブックを活用して計画を立てることが多くなっている。その際、目的とする観光スポットに関する概要やアクセスするための方法などの情報を集める。ガイドブックは、有名な観光スポットや流行りのものに焦点を当て、そのスポットの料金や見どころ、そして食べ物などが掲載されており、その土地に初めて観光で訪れるユーザや、普段から頻繁に観光旅行をしないユーザにとって満足を得られるような内容が書かれている。そのため、ガイドブックは普段から頻繁に観光旅行をするようなユーザや、その観光スポットに対してこだわりの観点を持つユーザの個別の嗜好まで考慮できない。

Web 上の情報に関して、観光スポットに関する総合的な情報が掲載された公式サイト、過去にそこを訪れた観光者による旅行記ブログサイト、そして、評価が記載されたレビューサイトなどの様々なサイトがある。しかしながら、種々のサイトが混在しており、膨大な情報から自分の嗜好に合う観光スポットの情報を得ることは簡単ではない。目的の情報を収集するためには手動で閲覧し取捨選択しなければならない。旅行計画を立つために、まず訪問したいエリアを決定する必要がある。その後、観光スポット検索サイトやガイドブックの観光スポット情報を調べて観光スポットを決定する。ユーザの意識決定の材料の 1 つとして、るるぶトラベル^{*1}、Tripadvisor^{*2} やじゃらん^{*3}などの観光スポット検索サイトがある。これらのサイトには特定の観光スポットを訪問したことのあるユーザがレビューを投稿し、観光スポットに関する豊富な情報が存在している。観光スポット検索サイト上で観光スポットに付与される情報の例^{*4}を図??に示す。なお、これらの情報のうちユーザレビュー以外の情報(概要文・タグ・評価点など)はメタデータと呼ばれる。

観光スポット検索サイトに掲載された情報を基に、利用者(ユーザ)は自身の嗜好や目的

^{*1} <https://www.rurubu.travel/>

^{*2} <https://www.tripadvisor.com/>

^{*3} <https://www.jalan.net/kankou/>

^{*4} <https://www.jalan.net/kankou/130000/?screenId=OUW2201>



図 1.1 レビューサイト上で観光スポットに付与される情報例

に合致した特徴を持つ対象観光スポットを検索する。多くの観光スポット検索サイトでは、対象観光スポットに付与されるメタデータを利用した以下のような検索方法が提供されている。

- ユーザが指定したタグを持つ対象観光スポット
- 評価点などの数値メタデータに基づく並び替え
- 入力単語(クエリ)を概要文中に含むオブジェクトの検索

しかし、これらのメタデータを用いた検索方法だけでは検索が困難な場合も存在する。たとえば、対象観光スポットを実際に利用することでユーザが得られる印象・感情・体験などを表す特徴に基づき対象観光スポットを検索したい場合などが該当する。また、訪問したいエリア内に数多くの観光スポットが存在している。ユーザは検索エリアに関する事前知識がないため、検索された多数の観光スポットからどの観光スポットのレビューを読むべきか効率的に判断することは困難で、自身のイメージから外れない観光スポットを見つけることは容易ではない。そこで、本研究では、さまざまな観光スポットを効果的に理解するためには、ユーザが訪問した体験のある観光スポットとを使って未訪問スポットを検索する手法を提案した。観光支援システムを構築するにあたり課題となる点は 2 つ説明する。

1. 観光スポット間の対応付け

ある観光スポットに対応する観光スポットを判定する手法として、レビューやカテゴリが類似したコンテンツであるという観点で判定する方法が一般的である。観光スポットで用いられる単語に対する従来の特徴量 TFIDF は、観光スポット内での重要な

単語を判定する際に用いられる。しかし、従来手法のTFIDFはユーザの観光スポット選択における意図や他スポットへの関心について考慮されていない。そのため、ある観光スポットが他の観光スポットと比較するとき、その観光スポットの特徴を表すことができない。そこで、あるスポットに対し、既に訪問した観光スポットと比較した相対的特徴ベクトルを作成し、相対的特徴ベクトル間の類似度に基づいて既訪問スポットと未訪問スポット間の対応付け手法を提案する。第3章で、より具体的な提案手法について述べる。

2. 観光スポットの対応関係を地図上に可視化

地理的情報の可視化を意味的な関係の可視化に取り込む情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的である。また、意味的な関係性の可視化として、グラフモデル上でオブジェクト間のつながりを関連度で表現することが多い。観光スポット間の意味的な対応関係を可視化するために、未訪問スポットは地図上で実在する座標に固定され、配置の自由があるのは既訪問スポットのみという制約がある。そこで、ある未訪問スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できるために、既訪問スポットと未訪問スポットの対応関係を可視化する手法を提案する。第4章で、より具体的な提案手法について述べる。

以下に本論文の構成を示す。まず、2章で本研究のアプローチについて述べる。ここでは観光支援システムのアルゴリズムと関連研究について述べる。この観光支援システムのための2つの観点に基づく各手法は3章と4章で述べる。3章では、ユーザの既訪問スポットの位置付けに基づく未訪問スポットの説明手法について述べる。4章では、地図上における未訪問スポットの説明向上のための観光スポットの対応付け可視化手法について述べる。5章では、各手法に対する評価実験結果を踏まえた提案手法全体に対する考察を行う。6章では、まとめと今後の課題を述べる。

第 2 章

本研究のアプローチ

2.1 観光支援システム

観光支援システムの 1 つに、求める観光スポット情報を発見するために観光スポットの理解をしやすくすることが考えられる。観光情報収集までの過程を効率化することで、ユーザの観光計画への敷居は下がり、訪れたことない観光スポットの理解の向上も望める。本節では、ユーザい今まで体験したことがありかつ気に入った観光スポットを利用し、訪問したいエリア内の観光スポットに対して関連付けすることで、手軽に観光情報を理解する観光支援システムを提案する。観光支援システムの具体的な流れを図 2.1 に沿って説明する。

まず、ユーザの体験履歴の中ていままで訪れたことがあるかつ気に入った観光スポットを入力する。次に、これから行ってみたい都道府県やエリアを入力する。そして、データベースで保存されている 2016 年 10 月までのじゃらんの観光スポットレビューデータを使って提案手法に適用することで、ユーザの未知な観光スポットに対する理解を支援するため、既に訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを説明するために対応付けを行う。

本研究における観光スポットの対応付けに関して、ユーザによって、その強調すべき特徴が変化する。たとえば、日本の有名な観光スポットを履歴に持つユーザにとっての「東京都庁舎展望台」と「金閣寺」について考える。このとき、「金閣寺」の相対的特徴は、お寺、金色、京都などである。「東京都庁舎展望台」の相対的特徴は、展望、夜景、建物、新宿などが考えられる。

一方、京都の寺院を履歴に持つユーザにとっての「金閣寺」と「清水寺」について考える。このとき、同じ「金閣寺」でもお寺や京都などは特徴にはならず、「金閣寺」の相対的特徴は、金色、金箔、輝きなどとなる。「清水寺」の相対的特徴は、舞台や一望などである。

対応付けを行ったあと、ユーザに未訪問スポットを説明するための観光スポット特徴語を抽出する。最後に、地図上で既訪問スポットと未訪問スポットの対応関係と特徴語を可視化することで、ユーザが未訪問スポットに対する理解を容易化することを目指す。

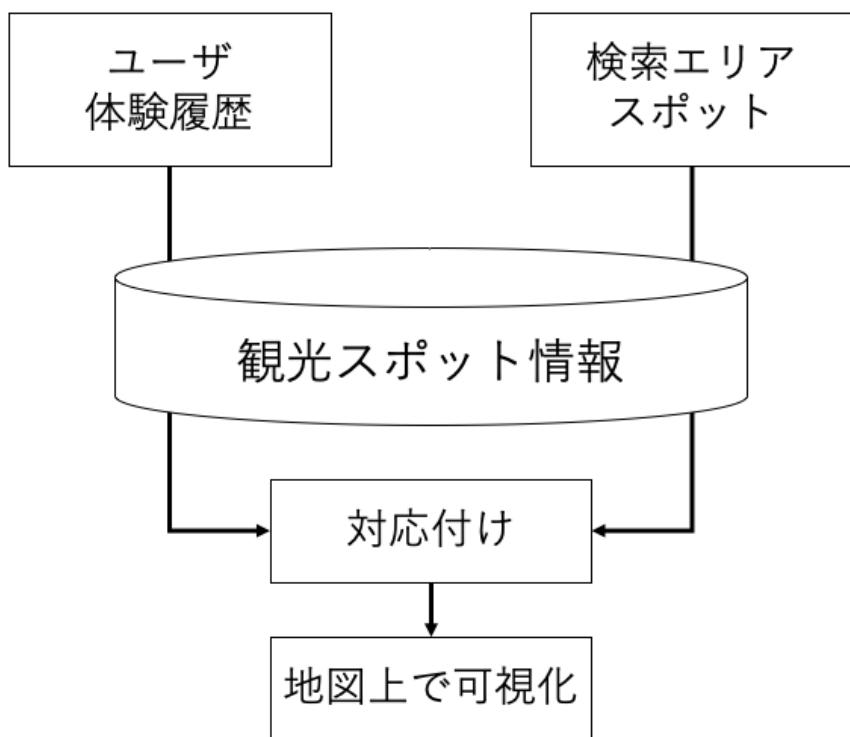


図 2.1 観光支援システムの流れ

2.2 関連研究

2.2.1 体験履歴による検索や推薦システムに関する研究

倉島ら [?] は、Flickr に投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案した。この手法では、ユーザの現在地から行きやすい場所とユーザの興味に合致した場所に移動しやすいと仮定し、行動モデルを生成している。ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用してユーザの行動モデルを生成している。

Cheng ら [?] は、自由に利用できるコミュニティ投稿の写真を活用して、パーソナライズされた旅行のおすすめに焦点を当て、特定のユーザプロファイルまたは属性を考慮し、パーソナライズされた旅行の推奨を行うことを提案した。

観光地検索するとき、松本ら [?] はクチコミから特徴語を抽出して利用する研究を行った。抽出対象を任意の名詞として、4種類の手法、TFIDF、ATF(Average Term Frequency)、ポアソン確率、エントロピーのうちどの手法が特徴語抽出に適しているのか検討を行った。また、抽出した特徴語を利用した検索支援システムを試作し、実験を通して特徴語提示の効

果を検証した。

嶋田ら [?] は Web から観光情報を抽出し、複数の特徴ベクトルから観光地間の類似性を評価することで、観光地を推薦するシステムを提案した。観光地の特徴ベクトルは、知恵袋・ブログ上での共起キーワードと時系列分布、知恵袋上でのカテゴリ構造、観光地周辺施設、地図画像から生成した。これらの特徴ベクトルから観光地間の類似度の測定を行い、類似度の高い観光地を推薦した。

廣島ら [?] は地理情報検索の際のクエリ入力支援として、提示する特徴語の抽出手法について研究を行った。この手法では、各ブログ記事から特徴語候補の抽出および地点の特定を行った。具体的には、特徴語の候補を Wikipedia の見出し語に限定し、ポアソン確率を用いて特徴語抽出を行った。また、地点の特定のための地名表現の抽出および緯度経度への変換の方法としては平野ら [?] の手法を用いた。

野守ら [?] は日本全国の観光地のクチコミデータを用いて、観光客が話題にする観光テーマを確率的に抽出し、そのテーマを軸として各観光地の特徴を定量的に評価した。また、クチコミのテキストデータにテキストマイニングを実行して表現を抽出し、観光地ごとにその表現の出現頻度を集計したクロス集計表に PLSA を実行することで、観光客のクチコミだけに基づいた観光テーマの抽出と観光地の特徴分析を行なった。

このように多くの研究はユーザに合致する観光スポットの抽出や、一般的な特徴の抽出に焦点をあてているが、本研究では、ユーザの選択性を高めるために、個人化した特徴を抽出する点では異なる。

上原ら [?] は Web 上に混在する観光情報からデータベースを構築し、ユーザのお気に入りの観光地を入力とし、複数の特徴ベクトルから観光地間の類似性を評価し推薦するシステムを構築している。

佃ら [?] は、スポットと属性値の関係の認知度を推定するための手法を提案している。2つのスポットの認知度と関連度といった指標を用いて、意外な共通点を発見する。

林ら [?] は、レビュー中の評価表現とその評価対象に基づき、ユーザが書いた映画レビューから他者が書いた映画レビューを推薦する手法を提案した。評価表現とその評価対象に基づく推薦を実現するために、レビュー中の「良い」や「良くない」などの肯定的または否定的な評価表現とその評価対象に着目した。

服部ら [?] はレビューを分析して、その価値観と繋がりの深い要素としてユーザの「こだわり」に着目し、その結果に基づくユーザモデリング手法について提案した。

従来のレビューを利用する手法では、クチコミを分析して、それぞれの嗜好や価値観を表す単語を重要視して、ユーザには単語レベルで提示されることが多い。単語レベルの提示で

は、レビューの文章の雰囲気を損なうため、ユーザの入力としては、体験したことかつ気に入った観光スポットの全レビューを利用すると考えた。

2.2.2 情報可視化に関する研究

本研究では、地理的な情報の可視化と意味的な関係の可視化に取り込む。地理的な情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的であり、本研究もそれに従っている。代表的な研究を以下に紹介する。櫻川ら[?]は、ソーシャルメディア上にアップロードされた写真のジオタグ情報と撮影時刻に基づいて写真の撮影者を分類する手法を提案した。分類された撮影者(観光者、在住者)ごとにホットスポットを可視化した。また、ジオタグ情報と撮影時刻以外にテキストタグを加えて、写真が撮影された地域で行われるイベントの穴場スポットを発見し、地図上に提示する手法を提案した[?]. 倉田ら[?]は、新しい観光情報サービスの形として「他の旅行者がどこで撮影したいか」というデータから観光ポテンシャルを可視化し、地図上に表示する方法を提案した。

意味的な関係性の可視化としては、グラフモデル上でオブジェクト間のつながりを関連度を表現することが多い。代表的な研究を以下に紹介する。Kitamuraら[?]は、一般的な物体認識を用いて、過去の個人旅行写真から推定したユーザの旅行の嗜好に基づき観光地を推薦する方法を提案した。物体認識システムを用いて、写真で撮った被写体情報のキーワードを取得し、グラフ視覚化技術によってキーワードの共起を表現した。また、グラフの視覚化技術に基づいて旅行写真付きのグラフを視覚化するユーザインターフェイスを紹介した。上村ら[?]は、ユーザが投稿したタグ付き画像を用いたファッショントylesの関係性の可視化手法を提案した。ファッショントylesは類似するタグを空間座標に固定することによって関係を表している。

本研究では、観光スポット間の意味的な対応関係を可視化するために、未訪問スポットは地図上で実在する座標に固定され、配置の自由があるのは既訪問スポットのみという制約がある。

第3章

ユーザの既訪問スポットの位置付けに基づく未訪問スポットの説明手法

本章では、ユーザの未知なスポットに対する理解を支援するため、既に訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを説明する手法を述べる。また、対応関係を表す特徴語抽出、観光スポットの対応付け、未訪問スポットの説明の有効性についての評価実験を行う。

3.1 概要

ユーザは検索エリアに関する事前知識がないため、どのスポットのレビューを読むべきか効率的に判断することは困難である。さまざまな観光スポットを効果的に理解するためには、ユーザが訪問した経験のあるスポットを使って未訪問スポットと類推できることが重要であると考えた。たとえば、日本に初めて訪れるフランス人旅行者に対し、未訪問スポットである東京の「表参道」をパリにおける「シャンゼリゼ通り」と表現すると理解しやすいであろう。この考え方は、ユーザが以前に経験した物事を現在の物事に適用する一種の類推である。既知の知識（ベースと呼ぶ）から概念（ターゲットと呼ぶ）を獲得するときに類推思考が働くとされる^[?]。構造の類似性には3種類あり、特徴の共有数で決まる「対象レベルの類似性」、ベースに存在する関係とターゲットに存在する関係の共有度に基づく「関係レベルの類似性」、および題の解法あるいは目標レベルでの類似性である「プラグマティックな類似性」とがある^{[?], [?]}。本研究では、類推の質を明示的に扱うため、「関係レベルの類似性」に近いと考えられる。

図3.1は、構築したプロトタイプシステムであり、訪問履歴として金閣寺等の京都のスポットを持つユーザが東京のエリアを検索した画面である。未訪問エリアにある迎賓館に対して金閣寺が豪華絢爛という観点で類似する様子を示している。このように、抽出したキーワードを提示することで、ユーザの未訪問スポットに対する理解の支援を目指す。



図 3.1 プロトタイプシステム

3.2 未訪問スポットと既訪問スポットの対応付け

未訪問スポットを理解しやすくするために、最も位置付けの近いユーザの既訪問スポットを対応付ける手法を提案する。まず、ユーザが既に訪問した複数個の観光スポットとこれから訪問したい観光スポットエリア情報を入力する。本手法では、観光スポットのユーザレビューを用いて特徴ベクトルを生成する。さらに、ユーザの観光地の位置付けを抽出するために、あるスポットに対し他の訪れたスポットと比較した相対的特徴ベクトルを算出する。同様に、未訪問スポットもエリア内の各スポットの特徴ベクトルを求める。また、未訪問エリアの各観光スポットの相対的特徴ベクトルを、そのエリアの他の観光スポットと比較して計算する。次に、相対的特徴ベクトルの類似度によって既訪問スポットを未訪問スポットと対応付ける。最後に、その関係性を説明するためのキーワードを抽出する。

3.2.1 観光スポットのユーザレビューを用いた相対的特徴ベクトル生成

2016年9月末までのじゃらんから得られたレビューデータを使用して、分散表現^[?]を用いて観光スポットの特徴ベクトルを作成する。このとき、観光スポット毎のレビューをまとめて1つの文書として扱う。本研究では、分散表現を計算するためにPythonのライブラリであるgensim^{*1}を利用する。学習方法として、Distributed Bag-of-Wordsを利用

*1 <https://radimrehurek.com/gensim/models/doc2vec.html>

して，各スポットの全レビューを使って 300 次元で作成したベクトルを使う。既訪問スポットや未訪問スポットのレビューベクトルは，形態素解析器である MeCab[?] に辞書「mecab-ipadic-NEologd」^{*2}用いて，分かち書きし原形にしたレビューを利用して作成する。

相対的特徴ベクトル $r_{state,i}$ を，式 3.1 として定義する。

$$r_{state,i} = s_i - \text{average}(\{s | s \in S_{state}, s \neq s_i\}) \quad (3.1)$$

相対的特徴ベクトルは，そのスポット自身の特徴ベクトルから他のスポットの特徴ベクトルの平均を引いた値によって得られる。 $S_{state} = \{s_1, s_2, \dots, s_n\}$ は，既訪問スポットもしくは未訪問スポットのベクトル集合である。 $state$ が f であれば，既訪問スポット集合を意味し， $state$ が u であれば，未訪問スポット集合を意味する。 s_i は集合 S_{state} 内の観光スポット i の特徴ベクトルを示している。

3.2.2 対応する観光スポットの決定と特徴語抽出

未訪問エリア内のスポットを既訪問スポットと対応付ける。具体的には，ある未訪問スポットの相対的特徴ベクトルに最も類似度が高い既訪問スポットを対応付ける。類似度計算には，コサイン尺度を用いる。ただし，類似度が閾値（本研究では 0.125）以下である場合は対応付けを行わない。

スポットの対応関係を示すだけでは，どのような点で対応するのかを理解するのは難しい。そこで，未訪問スポットと既訪問スポットの関係性を表すキーワードをユーザに提示する。しかし，分散表現である相対的特徴ベクトルから単語の特徴を得ることはできないため，他の方法を用いて単語を抽出する。

すべてのレビューを??節と同様に形態素解析器 MeCab によって単語に分割する。ただし，助詞，助動詞，連体詞，記号，ストップワードを削除している。

キーワード抽出手順について説明する。まず，TFIDF 法を使って対象となる既訪問スポットと未訪問スポットの特徴語と TFIDF 値を求める。IDF 値を算出する集合は S_{state} を用いる。

2 つのスポットに共通する特徴語の TFIDF 値の調和平均を用いて，対応付けしたスポットの説明可能なキーワードとして抽出する。各単語のスコアを式 3.2 によって定義する。

$$\text{score}(t, s_f, s_u) = \frac{2 \times \text{tfidf}(t, s_f) \times \text{tfidf}(t, s_u)}{\text{tfidf}(t, s_f) + \text{tfidf}(t, s_u)} \quad (3.2)$$

$\text{tfidf}(t, s_f)$ と $\text{tfidf}(t, s_u)$ は同じ単語に関する既訪問スポット s_f における TFIDF 値と未訪問スポット s_u における TFIDF 値を示している。単語スコアの上位 N 個の単語を説明情報

^{*2} <https://github.com/neologd/mecab-ipadic-neologd/>

表 3.1 既訪問スポット集合と未訪問スポット集合

既訪問スポット名	未訪問スポット名
浅草寺	東京ディズニーランド (R)
小田原城址公園	新宿御苑
伏見稻荷大社	東京スカイツリー
奈良公園	東京タワー大展望台
三島スカイウォーク	明治神宮

表 3.2 対応付け結果の例

未訪問スポット	既訪問スポット	特徴語
新宿御苑	小田原城址公園	お花見, 咲き誇る, 園内, 桜の時, のんびり, 手入れ, 自然, 遊具, ツツジ
東京スカイツリー	三島スカイウォーク	富士山, 摆れ, 高所恐怖症, 摆れる, 天井, 絶景, エレベーター, パノラマ, 展望デッキ, 昇る

としてユーザに提示する。

3.2.3 実行例

表 3.1 に示す, 既訪問スポット集合と未訪問スポット集合を用いて対応付けを行った結果を説明する。未訪問スポットは首都圏内を想定し 5 つのスポットを選んだ。表 3.2 は, ??節で述べた提案手法を用いて抽出した対応付いたスポットの特徴語を示している。

未訪問スポット「新宿御苑」に対し, 既訪問スポット「小田原城址公園」が対応付いた。他に候補として「奈良公園」が考えられるが, 「小田原城址公園」には花や遊具に対する記述が多く, 「奈良公園」には鹿や草に対する記述が多い。「新宿御苑」には花や遊具に対する記述が多いため, 「小田原城址公園」と適切に対応付けたと考える。

未訪問スポット「東京スカイツリー」に対し「三島スカイウォーク」が対応づけられた。どちらも眺めや高さに特徴があり関連付けられたと考えられる。また, 特徴語を見ると「富士山」が抽出されており, 見える景色を想起させるような対応付けを行っていると考える。

3.3 対応関係を表す特徴語抽出評価

3.3.1 実験内容

提案手法のうちの特徴語抽出部分について評価を行うため, 他のキーワード抽出手法と比較して評価を行う。以下の 3 つの方法を比較する。

- 算術平均法

表 3.3 対応関係を表す特徴語の実験結果

選択肢	算術平均	乗算	提案手法
1	28.28%	31.31%	29.29%
2	35.35%	31.31%	35.35%
3	10.10%	14.14%	12.12%
4	26.26%	23.23%	23.23%

- 乗算法
- 提案手法（調和平均）

キーワードを抽出するとき、算術平均法では、2つのスポットに共通する特徴語のスコアとして算術平均を計算する。抽出した単語のスコアを式 3.3 によって定義する。

$$score(t, s_f, s_u) = \frac{tfidf(t, s_f) + tfidf(t, s_u)}{2} \quad (3.3)$$

乗算法では、2つのスポットに共通する特徴語のスコアとして乗算を計算する。抽出した単語のスコアを式 3.4 によって定義する。

$$score(t, s_f, s_u) = tfidf(t, s_f) \times tfidf(t, s_u) \quad (3.4)$$

3.3.2 実験手順

クラウドソーシングのサービスである、 CrowdWorks^{*3}を利用して 23 人の被験者を集めた。まず、被験者は 4 から 10 個の自分が既に訪れたことのある観光スポットを入力した。次に、被験者は旅行などで行ったことがなく、これから訪れたい都道府県やエリアを入力した。

本実験では、それぞれの手法において、対応付けられた既訪問スポットと未訪問スポット、およびスポット同士の関係を説明する特徴語を最大で 5 つまで提示した。結果を評価するために、被験者は以下の 4 つの選択肢から 1 つを選んだ。

1. 2 つのスポットには関係性があり、キーワード^{*4}によって関係が明確になった。
2. 2 つのスポットの関係性があり、キーワードによって初めて気がついた。
3. 2 つのスポットには関係性があるが、キーワードは関係を表していない。
4. 2 つのスポットには関係性がない。

^{*3} <https://crowdworks.jp/>

^{*4} 関係を説明する特徴語のことであるが、被験者には単にキーワードと示した。

3.3.3 実験結果と考察

表 3.3 は、それぞれの手法における選択肢 1 から 4 それぞれの選択割合を示している。乗算法では、多くの被験者が選択肢 1 を選択している。このことから、乗算法は明らかな関係を表現する特徴語抽出ができることがわかった。算術平均法と提案手法では、選択肢 1 の割合が減少し、選択肢 2 の割合が増加している。このことから、算術平均法と提案手法は隠れた関係を表現する特徴語を見つけて被験者に提示することができると考えた。しかし、算術平均法では、選択肢 4 の割合が増加しているため、提案手法より無関係な特徴語を抽出しやすくなつたといえる。また、選択肢 1 と 2 を合計すると、提案手法が最も高くなつた。このことから、スポットの関連を説明する特徴語抽出手法としては提案手法が妥当といえる。

3.4 観光スポットの対応付けに関する評価

3.4.1 実験内容

特徴語抽出については??節の手法を用いる。提案手法のうちの対応付け部分について、以下の 3 つの手法を比較して評価を行う。

- メタデータ法
- 分散表現法
- 提案手法

メタデータ法は、観光スポット検索サイトでスポットを検索するためのメタデータであるカテゴリ、滞在時間、訪問時期の一致を用いる。まず、未訪問スポットに対し、すべてが一致する既訪問スポットを抽出する。抽出できない場合は、季節、滞在時間、カテゴリの順に条件を緩和する。既訪問スポットが複数ある場合は、レビュー数が最も多いスポットを選択する。メタデータの例を以下に示す。

- カテゴリ：神社・神宮・寺院、観光施設・名所巡り等
- 滞在時間：1 時間未満、1-2 時間等
- 訪問時期：1-12 月、春、夏、秋、冬

分散表現法では、各スポットの特徴として??節で作成した特徴ベクトルを使用する。すなわち、式 3.1 を用いずに、レビューから得た分散表現をそのまま用いる。

クラウドソーシングのサービスである、 CrowdWorks を利用して 24 人の被験者を集めた。実験手順は 3.3.2 節で説明した手順と同様である。

表 3.4 対応付けに関する実験結果

選択肢	メタデータ	分散表現	提案手法
1	41.30%	33.85%	29.36%
2	43.48%	47.69%	48.62%
3	2.17%	2.31%	2.75%
4	13.04%	16.15%	19.27%
対応付け数	46	130	109

3.4.2 実験結果と考察

対応付けしたスポットの合計数は 285 である。表 3.4 は、それぞれの方法における選択肢 1 から 5 それぞれの選択割合を示している。メタデータ法について、多くの被験者が選択肢 2 を選択している。このことから、メタデータ法は明らかな関係にあるスポット同士を関連付けることができることがわかった。他の方法と比較して、提案手法では、選択肢 1 の割合が減少し、選択肢 2 の割合が増加している。このことから、提案手法は隠れた関係を見つけて被験者に提示することができると考えた。しかし、選択肢 4 の割合も増加しているため、無関係なスポットを抽出しやすくなつたといえる。

選択肢 1 と 2 を合計すると、メタデータ法が最も高くなる。このことから、説明するためのスポットの対応付けの精度が最も高いといえる。しかし、メタデータ法は対応付けすることができるスポットの数が他の手法に比べ半分以下であり、未訪問スポットの多くに説明を付与することができない。

分散表現法と提案手法を詳細に考察するために、既訪問スポットの入力別に分析した。入力した既訪問スポットの半数以上が同一カテゴリのスポットであった被験者（表中の「同じ場合」）とそうではない被験者（表中の「異なる場合」）に分けて集計した。表 3.5 は選択肢 1 と選択肢 2 の評価の割合である。提案手法を使用すると、カテゴリが異なる場合において、分散表現法よりも選択肢 1 および選択肢 2 の割合を増加させることができる。相対的特徴ベクトルを用いることによって、カテゴリをこえて各スポットの特徴を求めることができるといえる。カテゴリが同じ場合の選択肢 1 は提案手法では、少なくなっている。これは、相対的特徴ベクトルとしては、カテゴリの特徴が打ち消されたベクトルが生成されるため、明らかな関係となる対応付けが減ったと考えられる。しかし、選択肢 2 が増えずに選択肢 4 が増え、無関係に見える対応付けを増やす結果となった。そのため、入力スポットの種類によって対応付け手法を切り変えることが有効と考えられる。

表 3.5 既訪問スポットのカテゴリの同異による評価割合

	選択肢	分散表現	提案手法
異なる場合	1	19.23%	21.10%
	2	24.62%	25.69%
同じ場合	1	14.62%	8.26%
	2	23.08%	22.94%

3.5 未訪問スポット説明の有効性評価

3.5.1 実験内容

提案手法の有効性を評価するために以下の 2 つのシステムを使って比較する。

- a. 未訪問スポット名のみを表示
- b. 未訪問スポット名、対応する入力スポット、対応を説明する特徴語を表示（提案手法）

被験者の入力は、これまでの実験と同様に、4 個から 10 個の自分が既に訪れたことのある観光スポットと訪問したいエリアである。ただし、エリアは 2 つ入力し、それぞれ別々にシステム a、システム b の入力として用いた。なお、システム a とシステム b はランダムな順番で実行される。システムを評価するために、被験者は以下の 2 つの設問について回答した。また、それらの選択理由について自由記述で答えた。

Q1. 表示されたスポットの詳細情報はどちらの方が分かりやすいかを選択してください。

Q2. 旅行の計画を立てる際にどちらの方が使いたいかを選択してください。

被験者はクラウドソーシングのサービスである、 CrowdWorks を利用して集めた 50 人を用いた。

3.5.2 実験結果と考察

Q1 に対する回答について、システム a とシステム b の結果はそれぞれ 12 件と 38 件となった。被験者の回答理由の例を表 3.6（システム a）および、表 3.7（システム b）に示す。また、旅行の計画を立てる際にどちらの方が使いたいかに対する回答について、システム a とシステム b の結果はそれぞれ 10 件と 40 件となった。被験者の回答より、得られた 2 つのシステムの特徴をまとめた。

表 3.6 システム a の方が良いと回答した被験者の自由記述

スポット名だけ書いてあるのでごちゃごちゃにならず、一目でどこに何があるか分かるからです。
スポットだけ表示されるところがシンプルで分かりやすかったから。
シンプルにそのエリアの観光スポットを知ることが出来る。

表 3.7 システム b の方が良いと回答した被験者の自由記述

自分が訪れた事のあるスポットと関連性があり連想しやすいから。
どのような場所か想像しやすいので、計画を立てるのにも使いたいと思ったから。
何と関連して表示された等詳細な情報が表示されわかりやすかった。

システム b では、キーワードや関連している情報が表示されており分かりやすいという特徴がある。これに対し、システム a では、スポット名のみが表示されているためシンプルで分かりやすいという特徴があるが、なぜそのスポットが検索されたのか分からぬなどの否定的な回答も多数あった。システムの表示情報が「シンプルな方が良い」と「詳細な情報が良い」の 2 つに大きく分けられているが、Q2 の回答より 80% の被験者はシステム b の方が使いたいと回答しており、提案手法の方が評価が高いといえる。また、システム b に関してスポットの写真を載せるとより分かりやすいなどの意見があり、画像による関連性の説明という拡張が考えられる。

第4章

地図上における未訪問スポットの説明性向上のための観光スポットの対応関係可視化手法

本章では、観光スポット間の対応関係を地図上に可視化する手法を提案する。また、プロトタイプシステムを構築し、既訪問スポットと未訪問スポットの説明性向上のための観光スポットの対応関係を地図上に可視化した効果を評価する実験を行う。

4.1 概要

近年、ユーザはWeb上の観光情報を活用して旅行計画を立てることが多くなっている。しかし、旅行は一般的に訪れたことがないスポットに行くことが多いため、観光情報を適切に理解することは困難である。そこで、訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを対応付けして、ユーザの未知なスポットに対する理解を支援することを考える。観光スポット間の対応関係を地図上に可視化する手法を提案する。本手法では、まず、スポットの特徴ベクトル間の類似度に基づいて既訪問スポットと未訪問スポットを対応付けし、その対応関係を説明するための特徴語を抽出する。次に、地図上で表示するための意味的な位置関係を決定する。具体的には、未訪問スポットからの距離と類似度によって存在確率を決定して、意味的な位置関係に基づいて観光スポットの対応関係をわかりやすくするために既訪問スポットの座標を算出する。また、プロトタイプシステムを構築し、既訪問スポットと未訪問スポットの説明性向上のための観光スポットの対応関係を地図上に可視化した効果を評価する実験を行う。

4.2 観光スポット間の対応関係を用いた可視化

未訪問スポットを理解しやすくするために、もっとも位置付けの近いユーザの既訪問スポットを対応付ける手法を提案した[?]. この手法では、ユーザが既に訪問した複数個の観

光スポットとこれから訪問したい観光エリアを入力とする。ユーザが入力した観光スポット集合のユーザレビューを用いてそれぞれのスポットの特徴ベクトルを生成する。同様に、未訪問スポットもエリア内の各スポットの特徴ベクトルを求める。次に、特徴ベクトルのコサイン類似度によって既訪問スポットを未訪問スポットと対応付ける。最後に、その関係性を説明するためのキーワードをTF-IDF特徴量と調和平均で抽出する。

ある未訪問スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できることが望ましい。ユーザは地図上で未訪問スポットを検索しているものとし、まず、地図上の実際にスポットが存在する座標に未訪問スポットを配置する。次に、既訪問スポットの位置を決定する。これは未訪問スポットとの対応関係によって変化し、未訪問スポットとの類似度が高いと近く、類似度が低いと遠くなるように配置することで、意味的な位置関係を表現する。既訪問スポットの座標は、式4.1によって求める。

$$Score(s_f, T) = \sum_{s_u \in S_u} w(s_u, s_f) \times P_{s_u, s_f}(d(s_u, T)) \quad (4.1)$$

S_u を未訪問エリア内のスポット集合とする。 $w(s_u, s_f)$ は、ある既訪問スポット s_f と未訪問スポット s_u の類似度が正であれば1.0、負であれば-0.5を返す関数である。 T は、候補の座標である。 $d(s_u, T)$ は、ある座標 T から未訪問スポット s_u のユークリッド距離である。また、 P_{s_u, s_f} は、平均 μ が0、標準偏差 σ が $(1 - |\cos(\text{vec}(s_u), \text{vec}(s_f))|) \times \alpha$ である正規分布を表す。 α は1である。 vec は未訪問スポットや既訪問スポットの特徴ベクトルを返す関数である。 $Score(s_f, T)$ がもっとも大きな座標 T を既訪問スポットの座標とする。図4.1は既訪問スポットの経度座標計算の例である。縦軸はスコアの値であり、横軸は経度である。4つの未訪問スポットを用いて、 $Score(s_f, T)$ (合計スコア)を計算している。その結果、135.06の座標がもっとも高くなる。緯度についても同様に行う。本稿では、近似として対象エリア内でランダムに200個の T を作り、それを用いた。

4.3 対応関係の可視化評価

4.3.1 実験内容

未訪問スポットの説明性向上のための観光スポットの対応関係可視化手法の評価を行うため、以下の3つの提示手法を使って比較する。図4.2、図4.3、図4.4は3つの提示手法それぞれのプロトタイプシステムとなっている。黒いピンは入力した既訪問スポット、赤いピンは未訪問エリア内のスポットを意味している。表示が重なることもあるため、黒いピンをクリックすると既訪問スポット名を吹き出しでも表示する。また、3つの提示手法において、類似度が「0.0」以下であれば、関係を表現するキーワードが「なし」と表示する。これは負

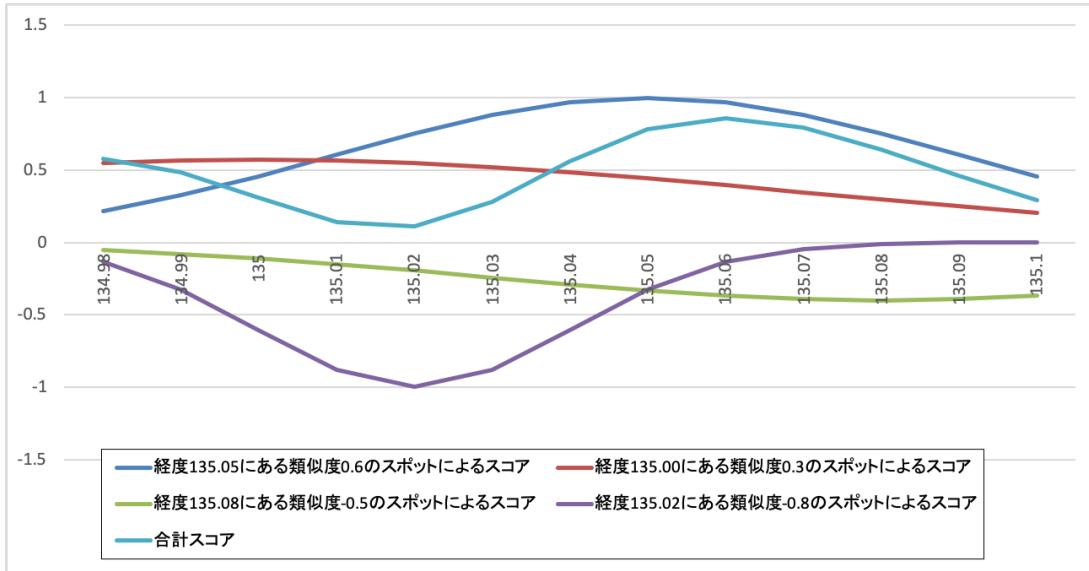


図 4.1 既訪問スポットの経度座標計算の例

の関係を表すキーワードは一般的に理解しにくいためである。

Position_Map(図 4.2)では、検索対象のスポットと既訪問スポットの関連度を全体的な位置関係で表現している。既訪問スポットは実際の座標ではなく、提案手法で決定した座標で表示される。赤いピンをクリックすると詳細情報が表示される。詳細情報は、未訪問スポット名、対応する既訪問スポット名、未訪問スポットと既訪問スポットの関係を表現するキーワードである。

Line_Map(図 4.3)では、位置関係による提示に加えて線の色を使って表現している。線の色は赤いピンと黒いピンの類似度を示している。線は緑色に近いと類似度が低く、赤色に近いと類似度が高いことを意味している。赤いピンをクリックすると詳細情報が表示される。詳細情報は、*Position_Map*と同じく、未訪問スポット名、対応する既訪問スポット名、未訪問スポットと既訪問スポットの関係を表現するキーワードである。

Table_Map(図 4.4)では、位置関係や線は用いずに、*Line_Map*で表現していたすべての情報を表形式でまとめて表現している。詳細情報は、現在選択している未訪問スポット名と、対応する既訪問スポット名、未訪問スポットと既訪問スポットの関係を表現するキーワード、類似度である。類似度の値は「-1.0 から 1.0」の間になる。「-1.0」はもっとも類似せず、「1.0」はもっとも類似することを意味している。

4.3.2 実験手順

クラウドソーシングのサービスである、CrowdWorksを利用して 75 人合計 120 件のデータを集めた。まず、被験者は 3 個から 9 個の自分が既に訪れたことのある、かつ気に入った



図 4.2 位置関係 (Position_Map)



図 4.3 位置関係 + 線 (Line_Map)



図 4.4 表形式 (Table_Map)

観光スポットを入力した。次に、被験者は旅行などで行ったことがなく、これから訪れたい都道府県やエリアを入力した。

本実験では、各提示手法において、検索スポットが重複しないように振り分け、3つの提示手法を連続して利用した。それぞれの提示手法において、被験者は検索スポットから行き先を決定した。このとき、表示順に影響をなくすために、6パターンの表示順を作成し、被験者はランダムに1つの表示順で実験を行う。実験を行った後、以下の項目について評価する。

- 各提示手法の行き先の決定しやすさ
- *Position_Map* と *Line_Map* どちらのほうが、訪問履歴との関係がわかりやすかったか
- *Line_Map* と *Table_Map* どちらのほうが、訪問履歴との関係がわかりやすかったか
- 各提示手法の訪問先決定までにかかった体感時間
- 各提示手法の訪問先決定までにかかった労力

表 4.1 行き先の決定しやすさの回答数割合

決定しやすさ	1	2	3	4	5	平均
<i>Position_Map</i>	5.0%	15.8%	15.0%	51.7%	12.5%	3.51
<i>Line_Map</i>	2.5%	12.5%	19.2%	50.0%	15.8%	3.64
<i>Table_Map</i>	3.3%	5.8%	35.0%	46.7%	9.2%	3.53

表 4.2 *Position_Map* と *Line_Map* の訪問履歴との関係の回答数割合

回答	回答数の割合
<i>Position_Map</i> がわかりやすかった	15.8%
<i>Position_Map</i> がややわかりやすかった	28.3%
どちらでもない	20.0%
<i>Line_Map</i> がややわかりやすかった	24.2%
<i>Line_Map</i> がわかりやすかった	11.7%

表 4.3 *Line_Map* と *Table_Map* の訪問履歴との関係の回答数割合

回答	回答数の割合
<i>Line_Map</i> がわかりやすかった	12.5%
<i>Line_Map</i> がややわかりやすかった	28.3%
どちらでもない	27.5%
<i>Table_Map</i> がややわかりやすかった	24.2%
<i>Table_Map</i> がわかりやすかった	7.5%

また、それぞれの手法について、自由記述で意見を書いてもらった。

4.3.3 実験結果と考察

表 4.1 は 3 つの提示手法それぞれの行き先の決定しやすさを 5 段階評価した回答数である。5 はとても決定しやすい、1 はとても決定しづらいを意味している。表 4.1 の平均値から *Line_Map* がもっとも決定しやすいことがわかった。*Line_Map* は *Position_Map* より決定しやすい理由として、*Line_Map* は位置情報提示に加えて、線を描くことによって、より既訪問スポットと未訪問スポットの対応関係を表現することができたためと考えられる。ただし、3 つの提示手法について *t* 検定を行った結果、それぞれ有意差があるとはいえないかった。

表 4.2 は *Position_Map* と *Table_Map* どちらのほうが、訪問履歴との関係がわかりやすかったかの回答数の割合となっている。表 4.2 から *Position_Map* がわかりやすいの割合が 44.2% に対して、*Line_Map* が 35.8% のため、*Position_Map* のほうがわかりやすいといえる。一方、表 4.1 では、*Position_Map* がもっとも決定しづらいといえる。理由として、

表 4.4 *Position_Map* と *Line_Map* の決定しやすさと関係わかりやすさについての回答数

決定のしやすさ	関係のわかりやすさ	回答数
<i>P</i> は <i>L</i> よりしやすい	<i>P</i> が <i>L</i> よりわかりやすい	18
<i>L</i> は <i>P</i> よりしやすい	<i>P</i> が <i>L</i> よりわかりやすい	3
<i>P</i> は <i>L</i> よりしやすい	<i>L</i> が <i>P</i> よりわかりやすい	1
<i>L</i> は <i>P</i> よりしやすい	<i>L</i> が <i>P</i> よりわかりやすい	23

表 4.5 訪問先決定までにかかった体感時間の回答数割合

	1	2	3	4	5	平均
<i>Position_Map</i>	10.0%	40.8%	30.0%	15.8%	3.3%	2.62
<i>Line_Map</i>	13.3%	35.8%	30.0%	15.8%	5.0%	2.63
<i>Table_Map</i>	9.2%	33.3%	35.8%	18.3%	3.3%	2.73

表 4.6 訪問先決定までにかかった労力の回答数割合

	1	2	3	4	5	平均
<i>Position_Map</i>	14.2%	35.8%	27.5%	16.7%	5.8%	2.64
<i>Line_Map</i>	15.8%	36.7%	30.0%	12.5%	5.0%	2.54
<i>Table_Map</i>	11.7%	33.3%	34.2%	15.8%	5.0%	2.69

Position_Map は他の 2 つの提示手法と比べると、決定しづらい回答数が多くなっている。また、「黒いピンと赤いピンの類似距離という関係性があまりよくわかりません」といった意見の被験者が多数おり、強く好まない被験者が一定数が存在するため、このような結果となつたと考える。位置関係を表現する際に、仮想的な地物の類似性を視覚的に表していることが明確にわかるインターフェースにする必要がある。また、表 4.3 は *Line_Map* と *Table_Map* の訪問履歴との関係のわかりやすさの回答数の割合となっている。表 4.3 から *Line_Map* がわかりやすいの割合が 40.4% に対して、*Table_Map* が 31.7% のため、*Line_Map* のほうがわかりやすい。

表 4.4 は同一の被験者の表 4.1 と表 4.2 の結果を比較し、*Position_Map* と *Line_Map* の行き先の決定しやすさと訪問履歴との関係のわかりやすさについての回答数となっている。表 4.1 に関しては、どちらに大きな値をつけたかを集計している。行き先決定のしやすさと訪問履歴の互いの影響に関して、決定しやすさと訪問履歴の評価が一致で回答した数はが異なると回答した数より多いため、互いに影響があることが考えられる。

表 4.5 は 3 つ提示手法それぞれの訪問先を決定するまでにかかった体感時間を 5 段階評価した回答数である。1 は訪問先決定するまで短い時間で、5 は長い時間で決定していることを意味している。表 4.6 は 3 つ提示手法それぞれの訪問先決定するまでにかかった労力を 5 段階評価した回答数である。1 は訪問先決定するまで少ない労力で、5 は多い労力で決定し

ていることを意味している。体感時間と労力の相関係数を計算すると 0.74 と強い正の相関があり、体感時間が短いと労力も少なく、体感時間が長いと労力も多くなる。訪問先決定までにかかった体感時間が短いと回答した数の割合は *Position_Map* が 50.8%, *Line_Map* が 49.2%, *Table_Map* が 42.5% となっている。よって、*Position_Map* がかかった体感時間がもっとも短く、*Table_Map* がもっとも長いことがわかった。訪問先決定までにかかった労力が少ないと回答した数の割合は *Position_Map* が 50.0%, *Line_Map* が 52.5%, *Table_Map* が 45.0% となっている。よって、*Line_Map* がかかった労力がもっとも少ない、*Table_Map* がもっとも多いことがわかった。以上から、*Line_Map* は体感時間が短くかつ労力も少なく訪問先を決定することができ、*Table_Map* はもっとも被験者にとって手間がかかっていることがわかった。対応関係を位置情報で表すことによって、表形式の時間と労力の問題点について改善することができるといえる。*Line_Map* は被験者に提示する文字の情報を少なくして、線の色と位置関係を使って既訪問スポットと未訪問スポット対応関係を見せるによって、他の 2 つの提示手法と比べると理解容易化することができたと考えられる。

第 5 章

考察

第 6 章

結論

謝辞

修士課程での研究全般に渡ってご指導賜りました北山大輔准教授に厚く御礼申し上げます。

お忙しい中貴重な時間を割いて副指導教員および副査を引き受けてくださり、ご指導賜りました三木良雄教授に感謝致します。

同様に、お忙しい中貴重な時間を割いて副査を引き受けてくださり、ご指導賜りました○○教授に感謝致します。

研究室配属からの3年間に渡って、切磋琢磨し心の支えにもなってくれた歴代のインタラクティブメディア研究室メンバ全員に感謝致します。

最後に、名前を書き切ることはできませんが、両親をはじめ、私のこれまでの人生を支えてくださった全ての方に心から感謝いたします。ありがとうございました。

本研究に関する発表

論文

- [1] 潘健太, 北山大輔, “説明性向上のためのユーザレビューを用いた観光スポットの対応付け手法,” 情報処理学会論文誌データベース, 第 84 号, 2019.

査読付き会議

- [2] Kenta Han, Daisuke Kitayama, “An Explanation Method of Unfamiliar Tourist Spots based on Roles of User’s Familiar Spots,” Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, Hong Kong, pp358-362, 13-15 March, 2019.

研究会

- [3] 潘健太, 北山大輔, “ユーザの既訪問スポットの位置付けに基づく未訪問スポットの説明手法,” 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), H7-2, 2019
- [4] 潘健太, 北山大輔, “地図上における未訪問スポットの説明性向上のための観光スポットの対応関係可視化手法,” 観光情報学会第 20 回研究発表会, 2019

その他の発表実績

研究会

- [1] 潘健太, 北山大輔, “ユーザのレビュー選択に基づく観光スポット検索手法,” 第 10 回 データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2018), H1-2, 2018
- [2] 潘健太, 北山大輔, “観光スポット検索のためのユーザのレビュー選択と特徴抽出に関する考察,” 電子情報通信学会技術研究報告, vol. 118, no. 107, DE2018-5, pp. 21-24, 2018

受賞

査読付き会議

- [1] Kenta Han, Daisuke Kitayama, “An Explanation Method of Unfamiliar Tourist Spots based on Roles of User’s Familiar Spots,” Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, Hong Kong, pp358-362, 13-15 March, 2019.